

03560.002919



SO.

6
BT
03-27-02

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

HIROAKI YOSHINO

Application No.: 09/976,098

Filed: October 15, 2001

For: SPEECH DATA RECORDING
APPARATUS AND METHOD
FOR SPEECH RECOGNITION
LEARNING

)
:
Examiner: UNASSIGNED

)
:
Group Art Unit: 2641

)
:
February 20, 2002

RECEIVED

FEB 25 2002

Technology Center 2600

Commissioner for Patents
Washington, D.C. 20231

SUBMISSION OF PRIORITY DOCUMENT

Sir:

In support of Applicant's claim for priority under 35 U.S.C. § 119, enclosed is a certified copy of the following foreign application:

2000-321435, filed October 20, 2000.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted

Attorney for Applicant
Gary M. Jacobs
Registration No. 28,861

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

GMJ/lmj
DC_MAIN 87528 v 1

09/976,098

CFG 2919 US

HIROAKI YOSHINO

"SPEECH DATA RECORDING APPARATUS AND METHOD FOR SPEECH
RECOGNITION LEARNING"

G.A.U. 26401PE

本 国 特 許 庁

JAPAN PATENT OFFICE

FEB 20 2002

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出 願 年 月 日

Date of Application:

2000年10月20日

出 願 番 号

Application Number:

特願2000-321435

出 願 人

Applicant(s):

キヤノン株式会社

RECEIVED

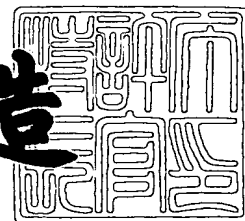
FEB 25 2002

Technology Center 2600

2001年11月 9日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3098730

【書類名】 特許願

【整理番号】 4323002

【提出日】 平成12年10月20日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/00

【発明の名称】 音声収録方法および音声収録装置および記憶媒体

【請求項の数】 18

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
社内

 【氏名】 吉野 宏昭

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
社内

 【氏名】 深田 俊明

【特許出願人】

 【識別番号】 000001007

 【氏名又は名称】 キヤノン株式会社

【代理人】

 【識別番号】 100076428

 【弁理士】

 【氏名又は名称】 大塚 康德

 【電話番号】 03-5276-3241

【選任した代理人】

 【識別番号】 100101306

 【弁理士】

 【氏名又は名称】 丸山 幸雄

 【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100115071

【弁理士】

【氏名又は名称】 大塚 康弘

【電話番号】 03-5276-3241

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0001010

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声収録方法および音声収録装置および記憶媒体

【特許請求の範囲】

【請求項 1】 音声認識処理における学習処理を行うために音声を収録する装置であって、

音声収録の対象となる文を表す収録文字列を格納する格納手段と、

学習のために入力された音声を認識して認識文字列を得る認識手段と、

前記認識文字列と前記収録文字列とを比較し、両者の一致度が所定レベルを越えるか否かを判定する判定手段と、

前記判定手段によって前記一致度が前記所定レベルを越えると判定された場合、当該入力された音声を学習のためのデータとして収録する収録手段とを備えることを特徴とする音声収録装置。

【請求項 2】 前記判定手段によって前記一致度が前記所定レベルを越えないと判定された場合に、当該文の音声による再入力を促す再入力手段を更に備える

ことを特徴とする請求項 1 に記載の音声収録装置。

【請求項 3】 前記判定手段は、前記収録文字列と前記認識文字列との間で DP マッチングを実行することによりその一致度を求める

ことを特徴とする請求項 1 に記載の音声収録装置。

【請求項 4】 前記判定手段による DP マッチングの実行の結果、前記収録文字列と前記認識文字列との間の不一致部分をユーザに提示する提示手段を更に備える

ことを特徴とする請求項 3 に記載の音声収録装置。

【請求項 5】 前記提示手段は、前記不一致部分の提示に際して、DP マッチングによって判定される挿入、脱落、置換の区別を可能とする提示を行う

ことを特徴とする請求項 4 に記載の音声収録装置。

【請求項 6】 前記提示手段は、前記収録文字列と前記認識文字列とを同時に画面上に表示するとともに、該収録文字列と該認識文字列の少なくともいずれかにおいて両者間の不一致部分もしくは一致部分の文字の属性を変更して表示す

る

ことを特徴とする請求項 3 に記載の音声収録装置。

【請求項 7】 前記提示手段は、前記収録文字列と前記認識文字列とを同時に画面上に表示するとともに、該収録文字列と該認識文字列の少なくともいずれかにおいて両者間の不一致部分もしくは一致部分の文字を点滅させて表示することを特徴とする請求項 3 に記載の音声収録装置。

【請求項 8】 音声認識処理における学習処理を行うために音声を収録する方法であって、

学習のために入力された音声を認識して認識文字列を得る認識工程と、

前記認識文字列と、音声収録の対象となる文を表す収録文字列とを比較し、両者の一致度が所定レベルを越えるか否かを判定する判定工程と、

前記判定工程によって前記一致度が前記所定レベルを越えると判定された場合、当該入力された音声を学習のためのデータとして収録する収録工程とを備えることを特徴とする音声収録方法。

【請求項 9】 前記判定工程によって前記一致度が前記所定レベルを越えないと判定された場合に、当該文の音声による再入力を促す再入力工程を更に備える

ことを特徴とする請求項 8 に記載の音声収録方法。

【請求項 10】 前記判定工程は、前記収録文字列と前記認識文字列との間で DP マッチングを実行することによりその一致度を求める

ことを特徴とする請求項 8 に記載の音声収録方法。

【請求項 11】 前記判定工程による DP マッチングの実行の結果、前記収録文字列と前記認識文字列との間の不一致部分をユーザに提示する提示工程を更に備える

ことを特徴とする請求項 10 に記載の音声収録方法。

【請求項 12】 前記提示工程は、前記不一致部分の提示に際して、DP マッチングによって判定される挿入、脱落、置換の区別を可能とする提示を行うことを特徴とする請求項 11 に記載の音声収録方法。

【請求項 13】 前記提示工程は、前記収録文字列と前記認識文字列とを同

時に画面上に表示するとともに、該収録文字列と該認識文字列の少なくともいずれかにおいて両者間の不一致部分もしくは一致部分の文字の属性を変更して表示する

ことを特徴とする請求項 1 0 に記載の音声収録方法。

【請求項 1 4】 前記提示工程は、前記収録文字列と前記認識文字列とを同時に画面上に表示するとともに、該収録文字列と該認識文字列の少なくともいずれかにおいて両者間の不一致部分もしくは一致部分の文字を点滅させて表示することを特徴とする請求項 1 0 に記載の音声収録方法。

【請求項 1 5】 請求項 1 乃至 7 のいずれかに記載の音声収録装置と、前記音声収録装置によって収録された音声データを用いて音声モデルの学習を行う学習手段とを備え、

前記学習手段で学習された音声モデルを用いて音声認識を実行することを特徴とする音声認識装置。

【請求項 1 6】 請求項 8 乃至 1 4 のいずれかに記載の音声収録方法によって音声进行収録する収録工程と、

収録された音声データを用いて音声モデルの学習を行う学習工程とを備え、

前記学習工程で学習された音声モデルを用いて音声認識を実行することを特徴とする音声認識方法。

【請求項 1 7】 請求項 8 乃至 1 4 のいずれかに記載の音声収録方法をコンピュータによって実行させるための制御プログラムを格納することを特徴とする記憶媒体。

【請求項 1 8】 請求項 1 6 に記載の音声認識方法をコンピュータによって実行させるための制御プログラムを格納することを特徴とする記憶媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、音声認識のための音声データ収録方法および装置、並びにこれを用いた音声認識装置および方法に関する。

【0 0 0 2】

【従来の技術】

一般に、音声認識は音響モデルおよび大量の音声データを収録した音声データベースが用いられる。このような音響モデルや音声データベースを構築するためには、大量の音声データを収録する必要がある。

【0003】

一般に音声認識は次の手順で行われる。まずマイクなどから入力された音声をAD変換して音声データを得る。ここで、マイクを通じて入力された音声には非音声区間も含まれるため、音声区間検出を行う。そして、音声区間検出がなされた音声データを音響分析することにより、ケプストラムなどの特徴量を計算する。音響分析により計算された音声データの特徴量から、音響モデル(HMM)に対する音響尤度の計算を行う。その後、言語探索を行って認識結果を得る。

【0004】

音響モデルとは、音素などの音声単位について、様々な発声者がどのように発声したかを記述したデータである。音声認識システムでは音声認識を開始する前処理としてユーザに少数の単語または文章を発声させ、その音声に基づいて音響モデルの修正(学習)を行うことで認識精度を高めている。また、音声認識の精度は音響モデル、ひいては大量の音声データを登録した音声データベースに負うところが大きく、その重要性はますます向上している。

【0005】

【発明が解決しようとする課題】

しかしながら、上記音響モデルの学習のためにユーザが発声した内容については、所定の単語または文章を間違いなく発声していると仮定するか、音声認識を用いて認識率からその単語または文章を正しく発声したかどうか判定するのみの簡単な正誤判定しか行っていない。また、音声データベースを構築するための、大量の音声データの収録・整備に要する時間とコストは非常に大きいため、この種の作業の効率化を図ることの必要性が高まっている。

【0006】

従って、本発明は、上記の課題に鑑みてなされたものであり、音声の切出し等が正しく収録されているか否かのチェックを自動で行い、発声ミスの少ない音声

データを効率よく収録可能とすることを目的とする。

【0007】

また、本発明の他の目的は、音声収録の際に、発生ミスユーザにわかりやすく提示することにより、音声収録に係る時間とコストを減らし、発生ミスの少ない音声データ収集を可能とすることにある。

【0008】

【課題を解決するための手段】

上記の目的を達成するための本発明による音声収録装置は、例えば以下の構成を備える。すなわち、

音声認識処理における学習処理を行うために音声を収録する装置であって、
音声収録の対象となる文を表す収録文字列を格納する格納手段と、
学習のために入力された音声を認識して認識文字列を得る認識手段と、
前記認識文字列と前記収録文字列とを比較し、両者の一致度が所定レベルを超えるか否かを判定する判定手段と、

前記判定手段によって前記一致度が前記所定レベルを超えると判定された場合、当該入力された音声を学習のためのデータとして収録する収録手段とを備える。

【0009】

また、好ましくは、前記判定手段は、前記収録文字列と前記認識文字列との間でDPマッチングを実行することによりその一致度を求め、

前記判定手段によるDPマッチングの実行の結果、前記収録文字列と前記認識文字列との間の不一致部分をユーザに提示する提示手段を更に備える。

【0010】

また、上記の目的を達成するための本発明による音声収録方法は、
音声認識処理における学習処理を行うために音声を収録する方法であって、
学習のために入力された音声を認識して認識文字列を得る認識工程と、
前記認識文字列と、音声収録の対象となる文を表す収録文字列とを比較し、両者の一致度が所定レベルを超えるか否かを判定する判定工程と、

前記判定工程によって前記一致度が前記所定レベルを超えると判定された場合

、当該入力された音声を学習のためのデータとして収録する収録工程とを備える。

【0011】

また、好ましくは、前記判定工程は、前記収録文字列と前記認識文字列との間でDPマッチングを実行することによりその一致度を求め、

前記判定工程によるDPマッチングの実行の結果、前記収録文字列と前記認識文字列との間の不一致部分をユーザに提示する提示工程を更に備える。

【0012】

また、本発明によれば、上記音声収録装置を備えた音声認識装置、音声収録方法を備えた音声認識方法が提供される。

更に、本発明によれば、これら音声収録方法或いは音声認識方法をコンピュータによって実現させるための制御プログラムを格納する記憶媒体が提供される。

【0013】

【発明の実施の形態】

以下、添付の図面を参照して、本発明の好適な実施形態を説明する。

【0014】

<第1の実施形態>

図1は、第1の実施形態による音声認識装置の音声収録に関わる機能構成を示すブロック図である。本実施形態の音声認識装置は、音声データベースの構築や音響モデルの学習のための音声収録に関して、以下の構成要素を備える。

【0015】

101は音声入力部であり、ユーザによる発声を電気信号に変化する。102はA/D変換部であり、音声入力部101からの音声信号をデジタルデータに変換する。103は表示部であり、音声収録すべき発声内容を表す発声リストを表示し、また、後述のマッチング部105によるマッチングの結果を表示する。104は音声認識部であり、A/D変換部102で得られる音声信号のデジタルデータに基づいて音声認識を行う。105はマッチング部であり、音声認識部104で得られた音声認識結果と発声リスト内容とのマッチングを行う。106は記憶部であり、マッチング部105によって正しく発声していると判別された音声

データを保存（収録）する。なお、以上の機能構成による音声収録処理は、図 3 のフローチャートにより更に詳細に後述する。

【 0 0 1 6 】

図 2 は、本実施形態による音声収録装置のハードウェア構成を示すブロック図である。図 2 において 2 0 1 はマイクであり、音声入力部 1 0 1 として機能する。2 0 2 は A / D 変換器であり、A / D 変換部 1 0 2 として機能し、マイク 2 0 1 からの音声信号をデジタルデータ（以下、音声データという）に変換する。2 0 3 は入力インターフェースであり、A / D 変換器 2 0 2 より得られる音声データをコンピュータバス 2 1 2 上に取り込む。

【 0 0 1 7 】

2 0 4 は CPU であり、当該音声認識装置の全体の制御を司るべく演算を行う。2 0 5 はメモリであり、CPU 2 0 4 からの参照が可能である。メモリ 2 0 5 には、音声認識ソフトウェア 2 0 6 が格納される。音声認識ソフトウェア 2 0 6 は音声収録処理のための制御プログラムを含み、CPU 2 0 4 はこれを実行することにより、表示部 1 0 3、音声認識部 1 0 4、マッチング部 1 0 5、記憶部 1 0 6 の各機能を実現する。また、メモリ 2 0 5 には、その音声認識および収録処理に必要な音響モデル 2 0 7、認識対象単語リスト 2 0 8 および言語モデル 2 0 9 が格納されている。更に、メモリ 2 0 5 には、音声収録のための発声内容を表す収録文リスト 2 1 2 が格納されている。

【 0 0 1 8 】

2 1 0 は出力インターフェースであり、コンピュータバス 2 1 2 と表示器 2 1 1 を接続する。2 1 1 は表示器であり、表示部 1 0 3 を構成するものであり、CPU 2 0 4 の制御により、例えば発声リスト 2 1 3 の内容や音声認識結果等を表示する。

【 0 0 1 9 】

次に、以上の構成を備えた本実施形態の音声認識装置による音声収録処理について説明する。図 3 は第 1 の実施形態による音声収録処理を説明するフローチャートである。

【 0 0 2 0 】

ステップS301では正しく発声されたかどうか判別するために認識結果と発声リストから求めた認識率を判定しきい値として設定する。ステップS302では、表示部103により、収録文リスト213に登録されている発声内容を表示器211上に表示することにより、音声収録のために発声すべき内容をユーザに提示する。ユーザがこの提示された収録文を読み上げると、音声入力部101を介してこの音声信号が取り込まれ、A/D変換102によって音声データに変換され、メモリ205に格納される（ステップS303）。ステップS304では、音声認識部104が、ステップS303で取り込んだ音声データについて音声認識処理を行い、その認識結果をメモリ205に格納する。

【0021】

次に、ステップS305では、マッチング部105がステップS304の認識結果とステップS302で提示した収録文との間のマッチングを行い、その認識率を求める。認識結果と提示文章との2つの文字列のマッチングにはDPマッチングを用いることができる。DPマッチングとは2つのパターン間で同じ文字同士が対応するように非線形に伸縮させる方法である。これにより2つのパターンの最小距離を求めることができる。また、不一致部分は、挿入、脱落、置換の3種類のいずれかのミスとして結果が得られる。なお、DPマッチングについては周知技術であるので、これ以上の説明は省略する。

【0022】

ステップS306では、ステップS305で求めた認識率がステップS301で設定した判定しきい値を越えたか否かに基づいて、正しく発声されたか否かを判定する。判定しきい値を越えない場合は、発声に誤りがあると判定し、ステップS307でDPマッチングの結果から発生ミス部分を表示器211上に提示し（表示部103）、ステップS303に戻る。こうして、当該提示文章に対する発声をユーザにやり直させる。

【0023】

一方、正しく発声されたと判定された場合は、ステップS308で、当該入力音声データを収録する。そして、ステップS309で、まだ発声すべき文章が収録文リスト213にあるか判断する。ここで、発声すべき収録文が残っていれば

ステップ S 3 1 0 へ進み、次の収録文を次の処理対象に設定してステップ S 3 0 2 に戻る。すべての収録文について発声が終わっていれば、ステップ S 3 1 1 に進み、本処理を終了する。

【0024】

ステップ S 3 0 7 における D P マッチングの結果の表示方法には種々の方法が考えられる。以下では、収録文が「五十五歳だってうれしいときはうれしいのだ」に対して、認識結果が「五十五歳でもうれしいときもあるのだよ」であった場合を例に挙げて、D P マッチングの認識結果表示方法の幾つかの例を示す。図 4 乃至図 6 は本実施形態による D P マッチングの認識結果の表示例を示す図である。

【0025】

図 4 の (a) は認識結果について、収録文と違う部分を背景色を変えて表示した例である。また、図 4 の (b) は、収録文について、認識結果と違う部分を背景色を変えて表示した例である。また、図 4 の (c) は、認識結果について、収録文と違う部分について、その発声ミスの内訳を「挿入」、「脱落」、「置換」の 3 種類に分けて、背景色を変えて表示した例である。すなわち、4 0 1、4 0 2 は置換が生じており、4 0 3、4 0 4 は脱落が生じており、4 0 5 には挿入が生じているので、「4 0 1、4 0 2」と「4 0 3、4 0 4」と「4 0 5」はそれぞれ異なる背景色で表示されることになる。

【0026】

上述の背景色の変更は、収録文或いは認識結果のいずれかの異なる部分について行われているが、両者の一致する部分の背景色を変更するようにしてもよい。この状態を図 4 の (d) に示す。なお、図 4 の (d) では、認識結果について一致する部分の背景色を変更しているが、収録文について認識結果と一致する部分の背景色を変更するようにしてもよい。

【0027】

なお、上記図 4 では文字列の背景色を変更することで一致部分或いは相違部分を示したが、文字の属性を変更することで一致部分或いは相違部分を示すようにしてもよい。図 5 の (a) は、認識結果について、収録文と異なる部分をフォン

トを変えて表示した例を示している。また、図5の(b)は認識結果について、収録文と違う部分を文字色を変えて表示した例を示している。文字の属性を変更する例としては、この他に、アンダーラインや影付きフォントへの変更等が挙げられる。また、図4(c)に示したように発声ミスの種類ごとにフォント属性を変えるようにしてもよい。

【0028】

以上、図4、図5では、異なる部分(或いは一致する部分)を静的に示したが、文字や背景の点滅等によって動的に示すようにしてもよい。図6(a)は、収録文と認識結果との相違する部分を点滅によって示す様子を示している。また、図6(b)は、収録文と認識結果との相違する部分の背景を点滅によって示す様子を示している。収録文と認識結果との一致する部分について上記表示を行うようにしてもよいことはいうまでもない。

【0029】

図7は認識結果のテキストと認識によって求められるワードグラフである。ワードグラフには認識された単語に対応する音声の開始位置と終了位置という情報が含まれている。そこで、ワードグラフから間違った単語の開始位置と終了位置を求め、マウスのクリックによって間違った単語を再生して音を聞く例である。

【0030】

以上のように、本実施形態によれば、学習のために入力された音声を認識して得られた認識文字列と収録文との比較によって両者の一致度が求められ、これに基づいて当該入力された音声を収録するか否かが決定されるので、発声ミスの少ない音声データを効率よく収録することが可能となる。

【0031】

また、本実施形態によれば、両者の一致度が所定レベルを越えないと判定された場合に、当該文の音声による再入力を促すので、容易に発声のやり直しをすることができ、効率の良い音声データ収録が可能となる。更に、両者の一致度はDPマッチングを用いて求められるので、置換、脱落、挿入等の不一致を正しく識別することができる。

【0032】

また、本実施形態によれば、上記DPマッチングの実行の結果によって得られる、収録文と認識文字列との間の不一致部分をユーザに提示するので、ユーザはどの部分に発声ミスがあったかを容易に把握できる。更に、不一致部分の提示に際しては、DPマッチングによって判定された挿入、脱落、置換の区別が可能なように提示が行われるので、より詳細に発声ミスを把握できることになる。

【 0 0 3 3 】

＜第2の実施形態＞

以上、第1の実施形態では、音響モデルの学習のための音声収録機能について説明した。第2の実施形態では、この音声収録機能を備えた音声認識装置について説明する。

【 0 0 3 4 】

図8は音声認識装置に本実施形態の音声収録装置を用いた音声認識装置の構成を示す図である。音声認識装置1301は、入力された音声に対してまず特徴抽出1302を行い特徴パラメータを抽出する。その後、音響モデル1304、言語モデル1305、発音辞書1306を用いて探索1303を行い認識結果を得るという流れで認識を行う。この図の例では認識精度を高めるため音響モデル1304を、話者にあわせて学習している。認識を始める前に、少数の学習サンプルを収録し、音響モデル1304の修正を行うのである。この学習サンプルの収録に際して、音声収録部1307が図3に示す音声収録処理を実行して音響モデル1304の学習を行う。

【 0 0 3 5 】

なお、本発明は、複数の機器（例えばホストコンピュータ、インタフェイス機器、リーダ、プリンタなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【 0 0 3 6 】

また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言

うまでもない。

【 0 0 3 7 】

この場合、記憶媒体から読出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【 0 0 3 8 】

プログラムコードを供給するための記憶媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

【 0 0 3 9 】

また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOS（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【 0 0 4 0 】

さらに、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【 0 0 4 1 】

【発明の効果】

以上説明したように、本発明によれば、音声の切出し等が正しく収録されているか否かのチェックを自動で行うので、発声ミスの少ない音声データを効率よく収録することが可能となる。

また、本発明によれば、音声収録の際に、発生ミスをユーザにわかりやすく提示することにより、音声収録に係る時間とコストを減らし、発声ミスの少ない音

声データ収集を効率よく収録することが可能となる。

【図面の簡単な説明】

【図 1】

第 1 の実施形態による音声認識装置の音声収録に関わる機能構成を示すブロック図である。

【図 2】

本実施形態による音声収録装置のハードウェア構成を示すブロック図である。

【図 3】

第 1 の実施形態による音声収録処理を説明するフローチャートである。

【図 4】

本実施形態による DP マッチングの認識結果の表示例を示す図である。

【図 5】

本実施形態による DP マッチングの認識結果の表示例を示す図である。

【図 6】

本実施形態による DP マッチングの認識結果の表示例を示す図である。

【図 7】

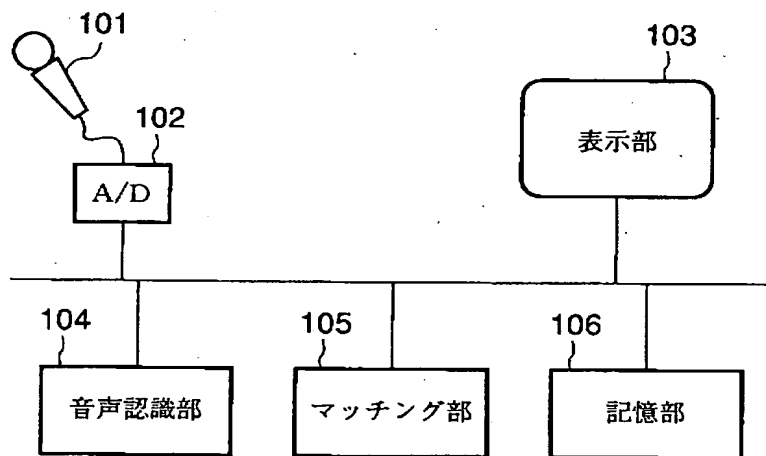
本実施形態による DP マッチングの認識結果の表示例を示す図である。

【図 8】

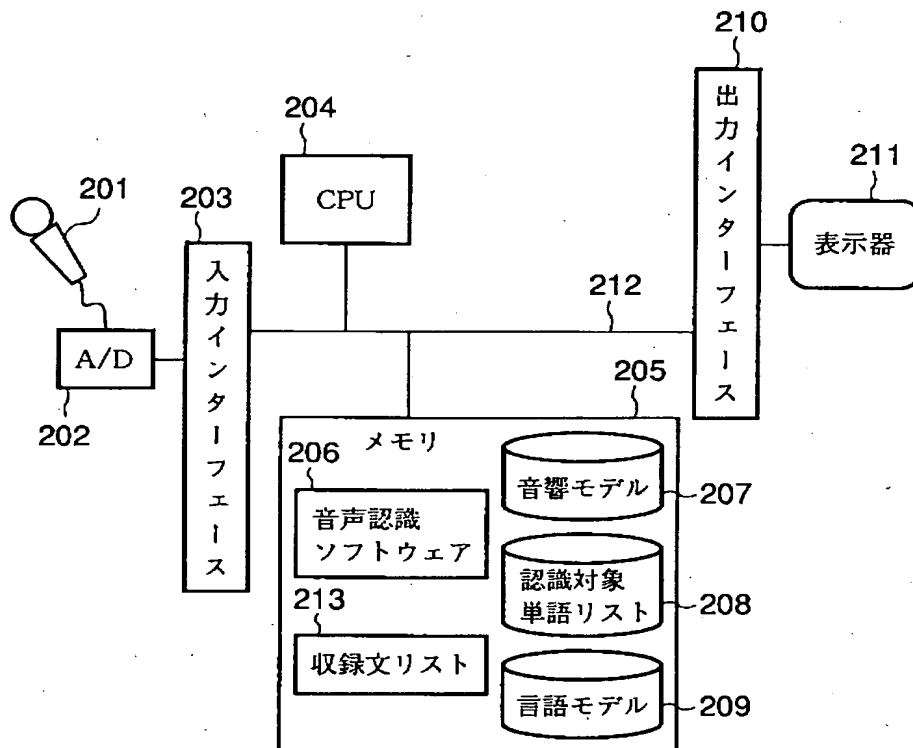
音声認識装置に本実施形態の音声収録装置を用いた音声認識装置の構成を示す図である。

【書類名】 図面

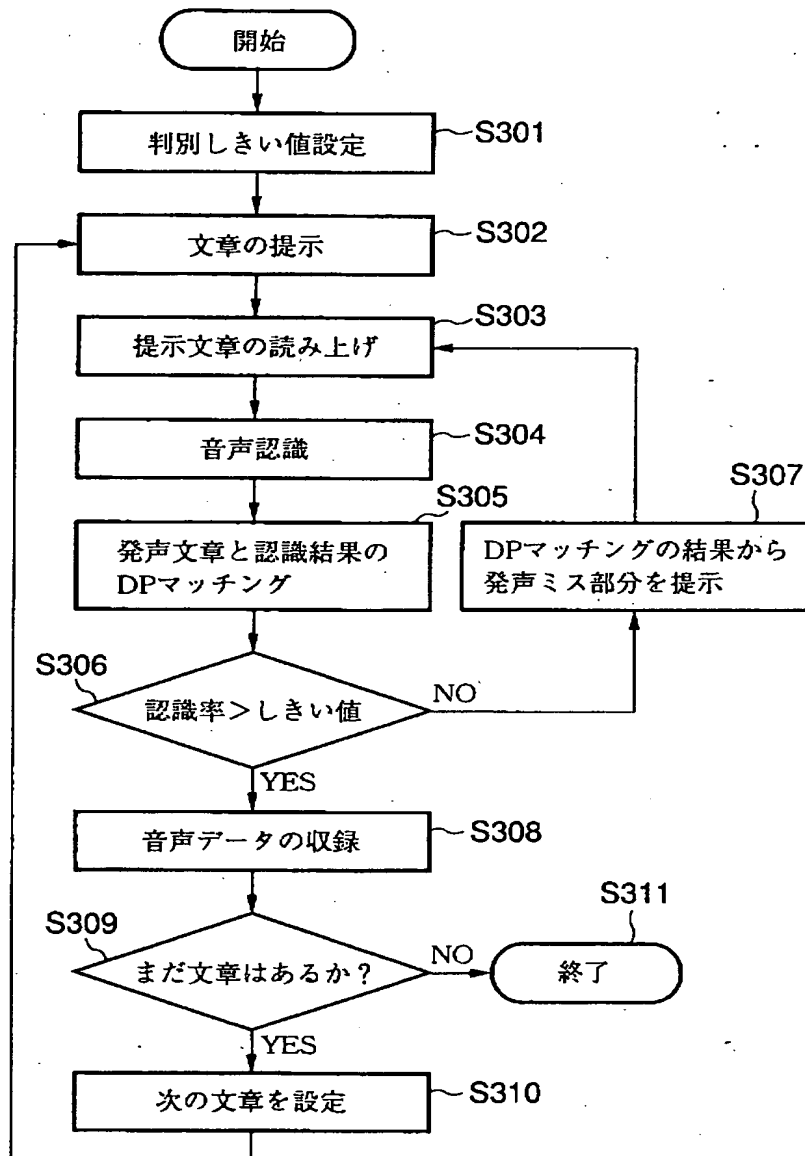
【図 1】



【図 2】


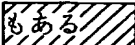



【図 3】


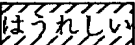



【図 4】






(a)

収録文 五十五歳だってうれしいときはうれしいのだ
 認識結果 五十五歳  うれしいとき  のだ  よ

(b)

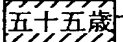
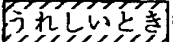
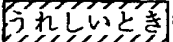



収録文 五十五歳  だってうれしいとき  うれしいのだ 
 認識結果 五十五歳でも うれしいときもある のだよ

(c)

収録文 五十五歳だってうれしいときはうれしいのだ
 認識結果 五十五歳   うれしいとき   のだ  よ

401 403 402 404 405

(d)

収録文 五十五歳だってうれしいときはうれしいのだ
 認識結果  五十五歳  でも  うれしいとき  もある  のだ  よ

【図5】

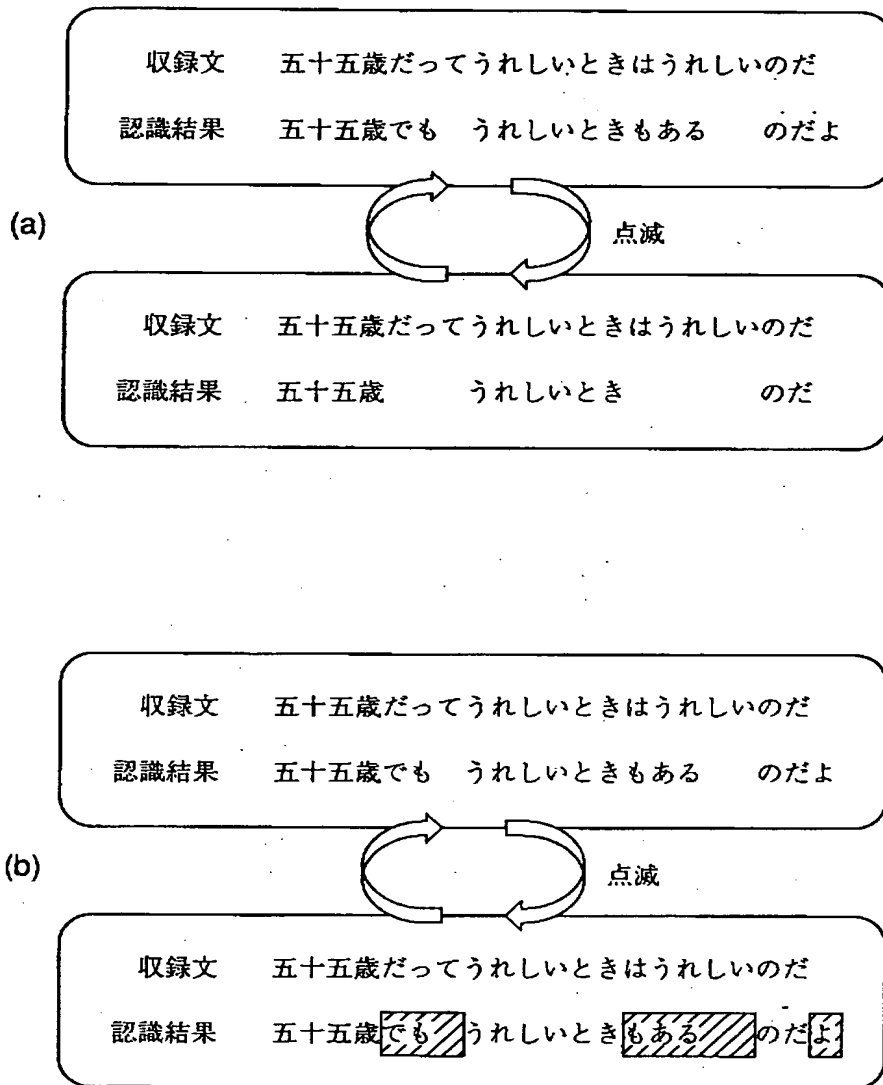
(a)

収録文 五十五歳だってうれしいときはうれしいのだ
認識結果 五十五歳でも うれしいときもある のだよ

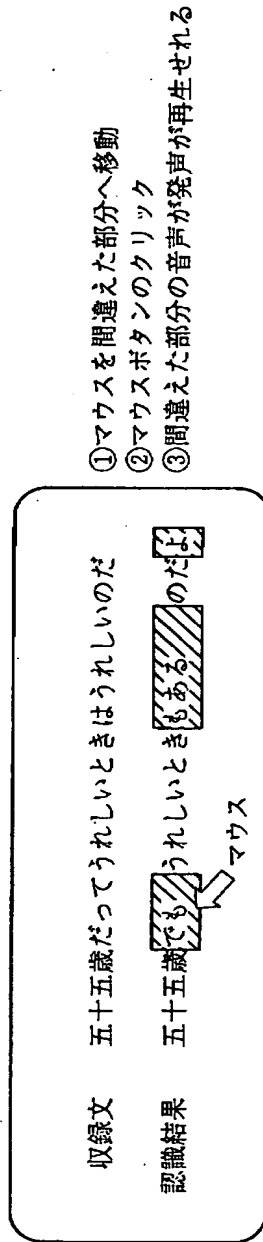
(b)

収録文 五十五歳だってうれしいときはうれしいのだ
認識結果 五十五歳でも うれしいときもある のだよ
黒 赤

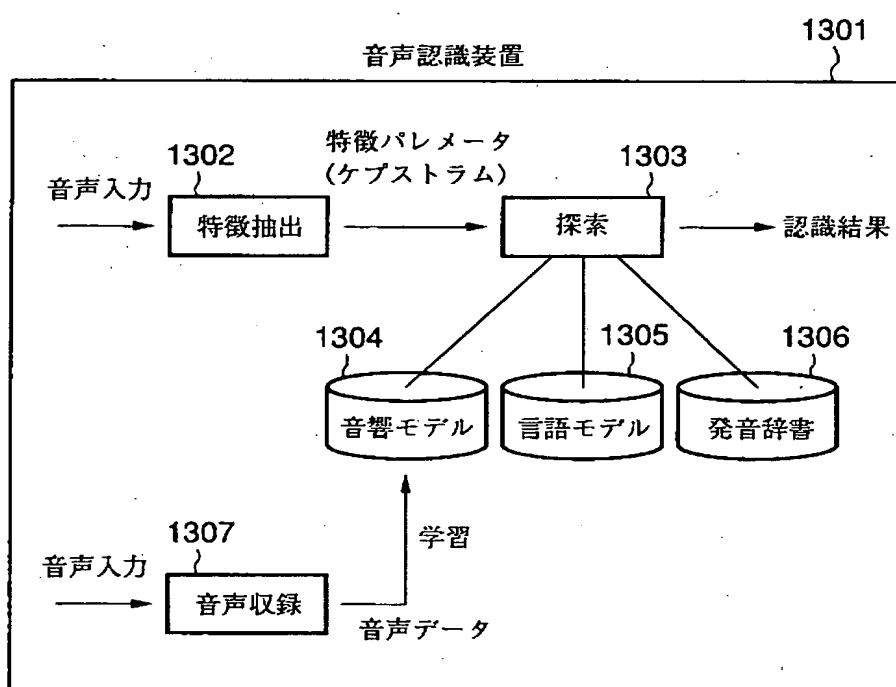
【図6】



【図 7】



【図 8】



【書類名】 要約書

【要約】

【課題】 音声の切出し等が正しく収録されているか否かのチェックを自動で行うことにより発声ミスの少ない音声データを効率よく収録する。

【解決手段】 音声認識処理における学習処理を行うために音声を収録するにおいて、学習のための収録文を提示し、これに対して入力された音声を認識して認識文字列を得る（ステップS302～S304）。得られた認識文字列と、上記提示した収録文とをDPマッチングによって比較し、両者の一致度が所定レベルを超えるか否かを判定する（ステップS305、S306）。ここで、一致度が所定レベルを超えると判定された場合は、当該入力された音声を学習のためのデータとして収録し（ステップS308）、そうでなければ、収録文と認識文字列との間の不一致部分をユーザに提示し（ステップS307）、当該収録分に対する再入力を促す（ステップS307、S303）。

【選択図】 図3

出 願 人 履 歴 情 報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日
[変更理由] 新規登録
住 所 東京都大田区下丸子3丁目30番2号
氏 名 キヤノン株式会社